UNCLASSIFIED

AD. **296 427**

*Reproduced
by the*

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

63-2-4

TECHNICAL NOTE NO. 5

under Contract
AF49(638)-1128

"Inductive Methods in Semantic
Analysis"

26 December 1962

THOMPSON RAMO WOOLDRIDGE INC.

RW DIVISION

CANOGA PARK, CALIFORNIA

Review of

Structure of Language and its Mathematical Aspects. Edited by
Roman Jakobson (Proceedings of Symposia in Applied Mathematics,
vol. XII). American Mathematical Society, Providence, R. I. , 1961.
vi + 279 pp. [1]

by Paul L. Garvin

Thompson Ramo Wooldridge Inc.
RW Division

0.   These are the Proceedings of the Twelfth Symposium in
Applied Mathematics, held in New York City, April 14-15, 1960.
The papers can roughly be grouped into the following five categories:
(1) Linguistics: Y. R. Chao, Graphic and Phonetic Aspects of Lin-
guistic and Mathematical Symbols (69-82), Gordon E. Peterson and
Frank Harary, Foundations in Phonemic Theory (139-65), H. A.
Gleason, Jr. , Genetic Relationship Among Languages (179-89);
(2) Mathematical and Logical Method: W. V. Quine, Logic as a
Source of Syntactical Insights (1-5), Nelson Goodman, Graphs for
Linguistics (51-55), Haskell B. Curry, Some Logical Aspects of
Grammatical Structure (56-68), Robert Abernathy, The Problem of
Linguistic Equivalence (95-8), Anthony G. Oettinger, Automatic
Syntactic Analysis and the Pushdown Store (104-29); (3) Pattern
Recognition: Murray Eden, On the Formalization of Handwriting;
(4) Models of Linguistic Structure: Noam Chomsky, On the Notion

"Rule of Grammar" (6-24), Hilary Putnam, Some Issues in the Theory of Grammar (25-42), Henry Hiż, Congrammaticality, Batteries of Transformations and Grammatical Categories (43-50), Morris Halle, On the Role of Simplicity in Linguistic Description (89-94), Hans G. Herzberger, The Joints of English (99-103), Victor H. Yngve, The Depth Hypothesis (130-8), Joachim Lambek, On the Calculus of Syntactic Types (166-78); (5) Probability and Information Theory: Benoit Mandelbrot, On the Theory of Word Frequencies and on Related Markovian Models of Discourse (190-219), Charles F. Hockett, Grammar for the Hearer (220-36), Rulon Wells, A Measure of Subjective Information (237-44), Roman Jakobson, Linguistics and Communications Theory (245-52). A summary of the discussion appears at the end of the volume (253-68).

The papers dealing with models of linguistic structure represent the central theme of the volume; also, their subject matter seems to have captured the imagination of linguists more than the other topics. This is the portion of the book, therefore, which I shall emphasize in this review. I shall take an outside view of the approach represented in it, by discussing it from the viewpoint of anthropological linguistics, rather than in terms of its own premises. In this connection, I should like to characterize the anthropological linguistic approach by two aims: as an immediate aim, the data-centered description of a great variety of languages, including many non-literate languages; as an ultimate aim, the classification of the relation of language to culture, both in a general theoretical sense, and in the particular historical sense of relating linguistic history to ethno-history.

The various models of structure are certainly not identical with each other, and their proponents are often in open disagreement (cf. R. B. Lees' Critique of Yngve in his comments to Hockett's paper, 266). Nevertheless, an outsider is able to spot certain recurrent conceptual motifs that can serve to characterize the "modeling" approach as a whole. These concepts will be presented in the form of a brief terminological glossary, organized in the following

2

five conceptual categories which are here presented in what I consider a decreasing order of generality:

1. models
2. the nature of grammar
3. types of grammars
4. the elements of language
5. rules.

Finally, the question of simplicity—the most significant value criterion of the modeling approach—will be commented on.

1. **Models.** Y. R. Chao in a recent paper cites 30 different "Synonyms and Characterizations of 'Model', "[2] showing the wide diversity of opinion and usage in regard to this term.

My own view is summed up in a previous paper:[3] "A Model can be a frame for the presentation of results or a frame for the acquisition of knowledge." The authors of the papers under review are primarily interested in the former, and the manner in which the results are obtained is often considered irrelevant. Most anthropological linguists would be more interested in models in the sense of a frame for the acquisition of knowledge. In both senses, a model roughly corresponds to the earlier "theoretical framework" or "conceptual framework."

[2] Y. R. Chao, Models in Linguistics and Models in General, Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress, edited by Ernest Nagel, Patrick Suppes and Alfred Tarski, pp. 563-4 (Stanford, Calif., 1962).

[3] Paul L. Garvin, From Model to Procedure, Proceedings of the National Symposium on Machine Translation, held at the University of California, Los Angeles, February 2-5, 1960, edited by H. P. Edmundson, p. 367 (Englewood Cliffs, N.J., 1961).

Models can be quantitative or qualitative. The ambition of the authors under review is to produce quantitative models of language; the technique of anthropological linguistics can, paraphrasing A. L. Kroeber,[4] be characterized as that of "qualitative contrast."

2. The Nature of Grammar. Chomsky's original definition that a grammar is "a device that generates all of the grammatical sequences of [a language] and none of the ungrammatical ones"[5] is acceptable to most model-oriented authors in the book. The following concepts are relevant to it.

2.1 Device (12, 130). This need not be a device in the literal sense but is primarily an "abstract logical" device, which is to say, "a set of rules" (6). It is closely related to an

Automaton (7, 13-4) which is an automatic device, both in the above abstract logical sense, and in the literal sense of an automatic machine or at least a computer program. Some language modelers are interested in the design of sentence-generating automata in the literal sense of computer programs;[6] others are satisfied with the verbal description of abstract logical devices.

---

[4] A. L. Kroeber, Ethnographic Interpretations, UCPAAE 47.198 (1957): "...it has mainly operated with qualitative precision of definition of pattern, sharpened in recent decades by successful use of the principle of contrast."

[5] Noam Chomsky, Syntactic Structures, p. 13 ('s Gravenhage, 1957).

[6] Victor H. Yngve, Random Generation of English Sentences, Proceedings of the First International Conference on Machine Translation of Languages and Applied Language Analysis (H. M. S. O., London, in press).

A Turing Machine (39 ff.) is an automaton with no restrictions on its storage or logical capacity, that is, with an infinite memory and circuitry. Such a device is by definition an engineering impossibility and therefore has to remain abstract. It is of great theoretical interest to mathematicians. Most modelers of language would like their grammar device not to be a Turing machine, since a Turing machine could simply contain in its infinite memory all conceivable sequences of a language and generate them from this store. This type of generation would not require rules, and would hence be uninteresting.

To most anthropological linguists, the grammar of a language is not a device but a part of the culture and/or its description.

2.2 Generate, generation (pp. 17 ff.). The terms produce, production (pp. 131 ff.) can be used as synonyms if the grammar device is considered in the literal, not merely the abstract sense.

The term "generation" is not necessarily understood in a literal sense (though it may be, if the device is considered in the literal sense). Rather, it is used to designate the capacity of a set of rules to characterize (42) or to give a complete specification of (6) the sequences to which these rules apply. A grammar which consists of such a set of rules is then considered a generative grammar, showing the generative paths (43) by which these sequences are said to have come about.

The comparable concept in anthropological linguistics is for a description to account for the data, both those actually at hand and, ideally, those yet to be collected. An ideal description in anthropological linguistics will give a complete account of the system of a language, just as an ideal generative grammar will give a complete specification.

The origin of the term "generate" is in mathematical parlance, where a particular definition is said to generate a set of entities.

5

But the linguistic uses of the term lack the precision of its use in mathematics.

2.3 Grammatical Sentences (40 ff., 167 ff.). This and not deviant utterances (25 ff.) is what the sequences generated by the device must be, in order for the device to be a grammar, i.e., an adequate model of the language. Lambek has "been unable to find agreement among modern linguists as to what constitutes a grammatical sentence" (167). To the anthropological linguist, grammatical sentences are the equivalent of culturally acceptable speech behavior.

2.4 "All and only" (7, 27) the grammatical sentences of a language. This is what a grammar device must be capable of accounting for. It is not sufficient to describe the forms contained in an actually collected corpus. The grammar device must completely specify all grammatical sentences, and hence at the same time reject all ungrammatical sentences. The difficulty of achieving this aim is discussed by Putnam (26 ff.); it is related to the problem of defining grammaticalness referred to further above.

Model-oriented theorists of language often misunderstand the use of a corpus in anthropological linguistics. The aim of the anthropological linguist is to infer from the corpus obtained in the course of field work the kind of description which will not be invalidated by additional data. This is achieved by basing a first approximation of a description on a given corpus, and then cross-checking it against additional field data. When a point of diminishing returns is reached in this process of cross-checking, it is assumed that no discrepancies can reasonably be expected, and hence that the description is valid and applies to the language in general. This is a less ambitious assertion than the one about "all and only," but it says very much the same, albeit more realistically.

3. <u>Types of Grammars</u>. The major dispute here seems to be whether or not a finite-state device (137) can be an adequate model of language. In non-model terminology, the question is whether or not speech behavior can be adequately described as a probabilistic succession of finite states. In linguistics, the debate seems to have been triggered by Hockett's description of the nature of speech communication in terms of certain aspects of probability theory.[7]

No anthropological linguist will deny that the probabilistic aspect of speech behavior is significant, although probabilistic calculations have so far been applied in historical linguistics (lexicostatistics), and not yet in grammatical description. The prerequisite for the application of probability theory is a knowledge of what the elements are that are to be considered in the calculation. Viewed in this light, matters of probability — including questions of finite state models — presuppose a grammar rather than constitute one.

4. <u>The Elements of Language</u>. Here is where the anthropological linguist finds himself on least familiar ground. Not only are a number of familiar terms either missing or given less prominence, but the key terms in the model-oriented discussion of language tend to be unfamiliar.

4.1 <u>Set</u>, as in <u>set of strings</u> (8) or <u>set of sentences</u> (6, 99). The term comes from mathematics: a set, in the mathematical sense, is "any collection of objects ... . A set is defined by specifying which objects are elements of the set. This may be done by stating some rule by which the elements can be identified, or by actually putting the elements on display."[8] (i.e., enumerating them).

---

[7] C. F. Hockett, <u>A Manual of Phonology</u>, pp. 6-11 (Baltimore, 1955).

[8] Irving Allen, <u>The New Mathematics</u>, p. 36 (A Mentor Book, New York, 1960).

As used by the modelers, the set of strings or sentences is defined neither by an identifying rule nor by an enumeration of the elements. The elements are identified simply by giving a verbal label.

4.2 <u>Sentences</u> (passim). These are primary givens in modeling terminology. Anthropological linguists might want to know how sentences are defined, or how they differ from utterances. This question is hinted at by Putnam when he cites Ziff's term "deviant utterance" (25) and asks why it is not called "sentence."

4.3 <u>Strings</u> (8 ff., 100 ff.) as in <u>strings of symbols</u> (8). These are simply sequences. The term "string" is used in mathematics as an informal notion denoting a sequence of symbols of any sort.

As used in the volume under review, relations are not included as part of these strings; they are expressed by the rules which are applied to the strings (see 5. below).

4.4 <u>Symbols</u> (passim). In modeling terminology, the smallest elements of a language are symbols. By symbols are meant, not symbols in the behavioral sense, but graphic symbols as the logician manipulates them. Language is said to consist of strings (or sequences, 42 ff.) of such symbols. Anthropological linguists would like to see greater emphasis placed on what the symbols stand for and how they were arrived at. Some of them would also prefer to restrict the use of symbols to phonemics and a few clearcut cases elsewhere in the grammar (such as some morphophonemic variations or simple relations in syntax), and to be content with verbal statements for the most part.

4.5 <u>Trees</u> (8 ff., 100 ff., 131 ff.) provide a graphic representation of the immediate-constituent analysis of strings. Trees have <u>nodes</u> (11, 51, 132) at the points where their branches (11)

separate, as well as at the ends of the branches. In a properly drawn tree, there is a symbol at each node.

Trees are in essence modern versions of the old parsing diagrams; their design depends on the interpretation of the constituent structure which they represent. The modelers are not interested in how this structure was discovered. Trees are considered a more efficient means of representation than the older diagrams such as the one shown in Hockett's text book,[9] and they are easier to draw.

4. 6   Nesting (111 ff.) or embedding (10 ff.) is the characteristic of a given construction of being enclosed within another. This is one of the properties of natural human language that makes its modeling difficult. Related to it is the question of depth (130 ff.), where the problem of the permitted number of nestings is considered. This appears to be a subject of some controversy; Yngve (ibid.) accepts George Miller's statement[10] that the maximum number of nestings, because of limitations on human memory, is seven. Others hold that nesting is potentially infinite.

Anthropological linguists would here ask a somewhat different question. Rather than assuming nesting as a universal, they would ask whether all languages allow nesting, and of those that do, to what extent they allow it.

5.   Rules (6 ff. , 89 ff.). These are the modeling equivalent of the anthropological linguist's statements. They differ from statements by being required to have a narrowly circumscribed form

---

[9]C. F. Hockett, A Course in Modern Linguistics, pp. 150 ff. (New York, 1958). Note, however, that in the volume under review he also presents a tree diagram (232).

[10]George A. Miller, Human Memory and the Storage of Information, IRE Transactions on Information Theory IT-2. 129-37 (1956).

(where statements can be more varied). An instance of this are the rewrite rules (8), of the form "rewrite X as Y," for generating constituent structures. They consist in rewriting symbols. (The statement equivalent would be: units are substituted for each other.) The restriction consists in requiring, for instance, that only one symbol be rewritten at a time. No such restriction applies to statements, and anthropological linguists would consider it arbitrary. The reason for this type of restriction is the emphasis on "describing the form of grammars,"[11] which implies the requirement of a very particular form. It is possible, of course, to envision any other particular form—the choice is then to be made on grounds of simplicity (see 6. below).

Recursiveness (22, 29) is considered an essential characteristic of rules, if they are to model language adequately. In the mathematics of discrete quantities, a recursive rule is one in which a quantity is not defined explicitly, but implicitly in terms of its predecessors.[12] In the volume under review, the term appears to be used in the less specific sense of iteration—it seems to characterize rules that allow consecutive re-application. There is no corresponding requirement in anthropological linguistics.

The order of application of rules (91 ff., 131 ff.) is essential in a generative model since it is one of the means of achieving simplicity (see below). The equivalent here is the order of statements in a description. This is just as significant to the anthropological linguist, but the proper order of statements is considered more a matter of good descriptive procedure in general than of simplicity in particular.

---

[11] Chomsky, op. cit. in fn. 5, p. 56.

[12] H. P. Edmundson, personal communication.

6.   Simplicity (38, 89 ff.). Simplicity is the evaluation criterion used to judge models of language. Halle (89 ff.) relates simplicity to the order of application of the rules; it is significant to note that he is able to discuss the concept in the area of phonology which most anthropological linguists would consider comparatively simple to begin with.

The criterion of simplicity presupposes that the question of factual correctness has already been resolved. This condition is not easily achieved in the case of little-known languages. Hence, the question of simplicity does not often arise; questions of consistency and particularly of exhaustiveness are much more likely to trouble the field worker or comparatist.

7.   Summary. We may note the significant difference in emphasis and interest between model-building and anthropological linguistics. The model-builder assumes that a reliable description is given and is interested in an "insightful" re-statement. The anthropological linguist, on the other hand, deals with data where a description is either absent or unreliable. He is therefore much more concerned with the original statement and with its reliability.